

Preparing Medical Imaging Data for Machine Learning.

Year: 2020 | Citations: 781 | Authors: M. Willemink, W. A. Koszek, Cailin Hardell, Jie Wu, D. Fleischmann

Abstract

Artificial intelligence (AI) continues to garner substantial interest in medical imaging. The potential applications are vast and include the entirety of the medical imaging life cycle from image creation to diagnosis to outcome prediction. The chief obstacles to development and clinical implementation of AI algorithms include availability of sufficiently large, curated, and representative training data that includes expert labeling (eg, annotations). Current supervised AI methods require a curation process for data to optimally train, validate, and test algorithms. Currently, most research groups and industry have limited data access based on small sample sizes from small geographic areas. In addition, the preparation of data is a costly and time-intensive process, the results of which are algorithms with limited utility and poor generalization. In this article, the authors describe fundamental steps for preparing medical imaging data in AI algorithm development, explain current limitations to data curation, and explore new approaches to address the problem of data availability.